Applied and Environmental Microbiology, Aug. 2010, p. 5269–5277 Vol. 76, No. 15
0099-2240/10/$12.00 doi:10.1128/AEM.00124-10

Host Association of *Campylobacter* Genotypes Transcends Geographic Variation▽†

Samuel K. Sheppard,[1] Frances Colles,[1] Judith Richardson,[2] Alison J. Cody,[1] Richard Elson,[2] Andrew Lawson,[2] Geraldine Brick,[2] Richard Meldrum,[3] Christine L. Little,[2] Robert J. Owen,[2] Martin C. J. Maiden,[1] and Noel D. McCarthy[1]*

Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, United Kingdom[1]; Gastrointestinal, Emerging and Zoonotic Infections Department, Centre for Infections, Health Protection Agency, London NW9 5HT, United Kingdom[2]; and Food, Water and Environmental Laboratory, Public Health Wales, Cardiff CF64 2XX, United Kingdom[3]

Received 18 January 2010/Accepted 25 May 2010

Genetic attribution of bacterial genotypes has become a major tool in the investigation of the epidemiology of campylobacteriosis and has implicated retail chicken meat as the major source of human infection in several countries. To investigate the robustness of this approach to the provenance of the reference data sets used, a collection of 742 *Campylobacter jejuni* and 261 *Campylobacter coli* isolates obtained from United Kingdom-sourced chicken meat was established and typed by multilocus sequence typing. Comparative analyses of the data with those from other isolates sourced from a variety of host animals and countries were undertaken by genetic attribution, genealogical, and population genetic approaches. The genotypes from the United Kingdom data set were highly diverse, yet structured into sequence types, clonal complexes, and genealogical groups very similar to those seen in chicken isolates from the Netherlands, the United States, and Senegal, but more distinct from isolates obtained from ruminant, swine, and wild bird sources. Assignment analyses consistently grouped isolates from different host animal sources regardless of geographical source; these associations were more robust than geographic associations across isolates from three continents. We conclude that, notwithstanding the high diversity of these pathogens, there is a strong signal of association of multilocus genotypes with particular hosts, which is greater than the geographic signal. These findings are consistent with local and international transmission of host-associated lineages among food animal species and provide a foundation for further improvements in genetic attribution.

Members of the genus *Campylobacter*, specifically *Campylobacter jejuni* and *Campylobacter coli*, are major causes of human morbidity worldwide and are the most common bacterial cause of gastroenteritis in industrialized countries (4). These bacteria are commonly found as apparently harmless members of the gut microbiota of many farmed and wild mammals and birds. This, together with the sporadic nature of most human disease, has contributed to the remaining uncertainty regarding the relative importance of different potential sources of human infection (34), inhibiting the implementation of effective public health interventions, which may have major economic consequences on intensive food production. Human infection with *C. jejuni* and *C. coli* has been epidemiologically linked to contact with pets and farm animals and to consumption of red meat, water, milk, and poultry (17, 20, 34).

The advent of multilocus sequence typing (MLST) for both *C. jejuni* and *C. coli* and its application to large and diverse isolate collections have enhanced understanding of the ecology (38) and epidemiology (10) of these important pathogens. It has been shown that there is substantial genetic differentiation between farmed ruminants and chickens (29) and even greater differentiation between farmed chickens and wild birds at the same farm site (5). Furthermore, MLST supports the application of population genetic attribution models (29, 44) to attribute human disease to host species of origin, based on reference data sets from a range of animal species and the environment. These studies (33, 37, 44) have confirmed observational studies (6, 18, 34, 42) which implicated the consumption of poultry, or food cross-contaminated from poultry, as an important source of human infection, accounting for 40 to 80% of cases. Although these genetic attribution studies confirm that a substantial proportion of infection comes from chicken, their main limitation was the restricted reference data sets available. The extent to which *Campylobacter* populations differ among host species and environmental niches and the importance of geographical and temporal effects are incompletely characterized. These are central to the success of genetic attribution studies.

The usefulness of the genetic attribution approach therefore motivates the establishment of larger and better-sampled reference data sets and highlights the need to understand the variation within these populations to improve estimates of the relative importance of host association to population structure compared to other effects. As well as supporting human disease attribution, this allows insights into bacterial ecology and evolution. The present study describes the genetic diversity and structure of a large representative collection of *C. jejuni* and *C. coli* isolates from retail poultry in the United Kingdom. This

* Corresponding author. Mailing address: Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, United Kingdom. Phone: 44-1865-281538. Fax: 44-1865-281275. E-mail: noel.mccarthy@zoo.ox.ac.uk.

† Supplemental material for this article may be found at http://aem.asm.org/.

▽ Published ahead of print on 4 June 2010.

collection was analyzed with published collections of isolates from a range of different host species and geographical areas to evaluate the impact of large-scale geographical distance on host-associated genetic differentiation.

MATERIALS AND METHODS

Retail poultry surveys. Isolates of *Campylobacter* from three retail poultry surveys (2001 to 2005), in which fresh and frozen chicken was sampled from retail outlets across the United Kingdom, were obtained from the archive held at the Health Protection Agency (HPA) Centre for Infections. A random stratified sample of 216 isolates was chosen from the archive of a United Kingdom-wide survey conducted between April and June 2001 (strata, 50% of isolates from England and 17% each from Wales, Scotland, and Northern Ireland, with 75% of isolates *C. jejuni* and 25% *C. coli* for each country). This survey has been reported fully elsewhere (15). The second survey (HPA—Coordinated Local Authority Sentinel Surveillance of Pathogens [CLASSP]) sampled raw whole chicken carcass samples (fresh or frozen) from butchers (30%) and large multiple retail chains (70%) in participating local authorities in England and Scotland, and 532 isolates were selected. The third study, a rolling surveillance study from Wales and Northern Ireland (30), used the same methods as the CLASSP study, and 255 isolates were chosen. The samples from CLASSP and the Wales and Northern Ireland studies used in the current study were from 2004 and 2005.

Microbiology. *Campylobacter* isolates were retrieved from the archive where they had been stored at −80°C in Microbank cryovials (Prolab PL1605/G) and subcultured on Columbia blood agar (CBA). Plates were incubated for 48 h in a MACS-VA500 microaerophilic workstation (Don Whitley Scientific, Ltd.) under microaerobic conditions (5% CO_2, 5% O_2, 3% H_2, and 87% N_2) at 37°C. All microbiology procedures were performed according to standards required for accreditation by Clinical Pathology Accreditation (United Kingdom). Original isolates were identified as either *C. jejuni* or *C. coli* based on hippurate hydrolysis in the 2001 survey (3), and for the two subsequent surveys, a molecular method based on 5′-nuclease (TaqMan) real-time PCR assays to detect genes specific for *C. jejuni* (*mapA*) and *C. coli* (*ceuE*) was used (2). Lysates were prepared as a cell suspension of each culture made in 125 μl phosphate-buffered saline or in water (Sigma Aldrich, United Kingdom) in a 0.2-ml PCR tube. Suspensions were vortexed and transferred to a heat block at 100°C for 5 min. This killed-cell suspension was clarified by centrifugation at 9,000 × *g* for 10 min and stored at −20°C.

PCR, sequencing, and bioinformatics. DNA template arrays were created in 96-well Thermo-fast, polypropylene plates (Abgene, United Kingdom), and seven-locus MLST was carried out by standard methods using published primers (9, 32). Each 25-μl PCR mixture comprised molecular-grade water (Sigma-Aldrich, United Kingdom), 2.5 μl 10× PCR buffer (Qiagen, Ltd.), 0.25 μM (each) forward and reverse primers, 0.2 mM deoxynucleoside triphosphate (dNTP) mix (Invitrogen, Ltd.), 0.025 U/μl (0.125 μl) *Taq* polymerase (Qiagen, Ltd.), and 2 μl of template DNA. The PCR thermal cycle began with a 15-min denaturation step at 95°C, followed by 35 cycles of 94°C for 30 s, 50°C for 30 s, and 72°C for 1 min, with a final extension at 72°C for 5 min. Five microliters of PCR products was visualized by UV transillumination following electrophoresis at 200 V (10 min) on a 1% (wt/vol) agarose gel in 1× Tris-borate-EDTA (TBE) buffer (1 mM EDTA, 40 mM Tris-acetate). The amplification products were purified by precipitation with 20% polyethylene glycol (PEG)–2.5 M NaCl (12) and stored at −20°C. Nucleotide sequencing PCRs were performed in both directions, using the same primers (forward or reverse) diluted (1:15) in water. Reactions were carried out in 10-μl volumes containing 2 μl of PEG-precipitated DNA resuspended in water, 1.0 μl 5× buffer, 0.02 μl BigDye Terminator v3.1 mix (Applied Biosystems, United Kingdom), and 0.25 μM either the forward or the reverse primer. The cycling parameters were as follows: 30 cycles of 96°C for 10 s, 50°C for 5 s, and 60°C for 2 min. Unincorporated dye terminators were removed by precipitation of the termination products with 95% ethanol, and the reaction products were separated and detected with an ABI Prism 3730 automated DNA sequencer (Applied Biosyststems, United Kingdom). Forward and reverse sequences were assembled from the resultant chromatograms using the Staden suite of computer programs (40). The consensus sequences were queried against the *Campylobacter* database to give an allele number. The combination of alleles for the seven housekeeping genes gave the sequence type (ST), with STs assigned into genetically related clonal complexes, based on sharing four or more alleles with the central genotype using the PubMLST database based on mlstdbNet software (19).

Published data. Reference data sets were assembled from published studies in Europe, Africa, and the United States to provide comparison data sets for analysis (9–11, 16, 21, 23, 24, 26, 28, 29, 31, 36, 37, 41). For *C. jejuni*, these comprised 422 cattle isolates, 431 chicken isolates, 188 sheep isolates, 91 environmental isolates, and 160 wild bird isolates. For *C. coli*, this comprised 98 cattle isolates, 253 chicken isolates, 54 sheep isolates, 380 pig isolates, and 110 turkey isolates. Isolate source data and ST information for the data used in this study, including data sets obtained from published sources, are included in Tables S1 and S2 in the supplemental material for *C. jejuni* and *C. coli*, respectively.

Genealogical analysis. Genealogies were reconstructed using ClonalFrame software (7). This model-based approach to determining bacterial microevolution distinguishes point mutations from imported chromosomal recombination events—the source of the majority of allelic polymorphisms (43)—allowing more accurate calculation of clonal relationships (7, 8, 38) in *Campylobacter* species, where recombination is substantial (14, 43). A 75% consensus cutoff was used with branches identified in 75% or more program sample iterations being accepted into the trees. The trees presented are consensus trees of 6 ClonalFrame analyses under these conditions, each with a 10,000 burn-in and 50,000 iterations.

Assignment methods. Assignment of isolates to host species based on MLST allelic profile was performed to assess the relationship of the isolate populations in this study with other published data sets of isolates from chicken and other host species. In each analysis, a set of isolates from one host population and geographical area was assigned between several other isolate collections from a range of hosts and/or areas, described below as the reference data set. The assignment is probabilistic based on the allele frequencies in the populations in the reference data set for each of the seven MLST loci. This analysis was performed using Structure, a Bayesian model-based clustering method designed to infer population structure and assign individuals to populations using multilocus genotype data (35). Differences in genotype frequencies between populations in the reference data set allow probabilistic assignment of isolates to these populations, even if there is some sharing of genotypes between the reference data set populations. The method is not biased by sample size differences, but it does need substantial sample sizes for each population in the reference data set to allow accurate assignment. The largest available data sets therefore were used in the reference data set. MLST allelic profile data were analyzed as allelic profiles. Analyses were performed with 10,000 iterations following a 1,000-iteration burn-in using the no-admixture model of Structure, and the isolate collection to be assigned was distinguished from the reference data set populations using the "usepopinfo" flag.

For *C. jejuni*, two assignment analyses were performed. In one, 53 chicken-derived isolates from the Netherlands (29) and in the other 46 chicken-derived isolates from Senegal (24) were assigned to a reference set. In both analyses, this comprised 742 isolates from the United Kingdom retail poultry survey described here, 590 from United Kingdom ruminants (cattle and sheep), 160 from United Kingdom wild birds, and 91 from United Kingdom environmental samples. The details of the samples from published sources are given in the original studies (10, 16, 24, 28, 29, 37). For *C. coli*, three comparisons were performed. First, 129 isolates from chicken in the United States (31) and 36 from chicken in Senegal (23) were assigned to a reference data set including 261 isolates from United Kingdom retail poultry, 110 from U.S. turkeys, 285 from U.S. pigs, and 60 from U.S. cattle. The details of the U.S. samples are described in the original publications (23, 31). The third comparison assigned the 285 isolates from U.S. pigs to a reference data set with isolates from U.S. chickens, turkeys, and cattle but with 95 pigs from published studies in the United Kingdom and Denmark (23, 26, 31, 36, 37).

By structuring the comparisons in this way, the assignment of Senegalese and Netherlands chicken isolates was independent of geography since the reference data sets were from other countries. We used non-U.S. chicken isolates but U.S. isolates from other species in the reference data sets to assign U.S. chicken isolates. This means that any host association observed represents the excess of host association over geographic association. This was done similarly for the assignment of U.S. pig isolates.

Analysis of molecular variation and F_{ST}. As an alternative approach, analysis of molecular variation was undertaken to compare the effects of geography and host species where at least two or more host species were sampled in each of two or more locations. These data were available for *C. jejuni* from chickens and cattle in the United Kingdom (1,040 chicken and 402 cattle isolates) and Finland (33 chicken and 20 cattle isolates) and for *C. coli* from cattle, chickens, and pigs from the United States (60 cattle, 129 chicken, and 285 pig isolates) and Europe (38 cattle, 349 chicken, and 95 pig isolates). Isolates were placed in populations for each host (within one geographical area), and these populations were grouped according to their geography. The analysis was undertaken with Arlequin software (13) using default values to attribute the percentage of molecular variation that is associated with group membership and population membership

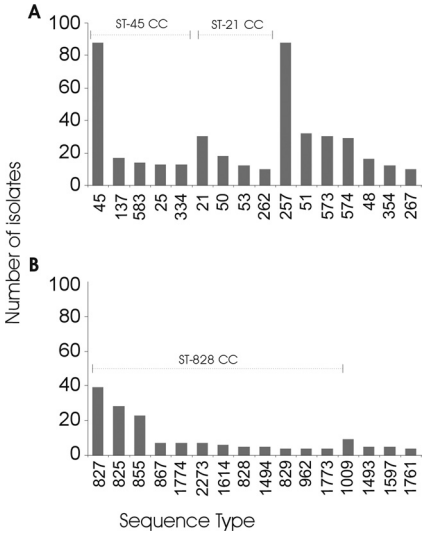

FIG. 1. The most common *C. jejuni* (A) and *C. coli* (B) genotypes in United Kingdom retail poultry. Frequencies are determined for sequence types (STs) derived from seven-locus MLST, and the clonal complex (CC) is indicated where several of these STs are from a single clonal complex.

within groups and within individual populations. This provided an alternative approach to the separation of the effects of host and geography.

Neighbor-joining F_{ST} (fixation index) trees were drawn with Mega software (25) using a matrix of F_{ST} values calculated in Arlequin (13). The F_{ST} values were calculated using 3,309-bp concatenated nucleotide sequences of MLST genes.

Testing host-geographic lineage association. The host association of STs was mapped on ClonalFrame trees to explore the nature of this association. The hypothesis that recently expanded host-associated or geographically restricted lineages exist was tested as follows. Lineages containing two or more STs that shared a common ancestor at ≤0.07 coalescent units were defined as a clade. A score for the association of a clade with a particular host or area was calculated as follows. Each ST was assigned to one host species (or area) if isolated only from that species (or area) and equally between the two if isolated from both. A score was then calculated for each clade as the number of STs isolated from one host species (or area) minus the number isolated from the other species (or area) in that clade. Finally the positive values of these scores from each clade were added together as a measure of the extent of association of clade with host species (or area). Permuted scores were obtained by randomly assigning each ST among the clades and then calculating the score as described above. This was repeated 10,000 times for each analysis to allow a comparison of the observed score to the range of scores that might arise randomly.

Nucleotide sequence accession number. Data from this study have been submitted to the PubMLST database (http://pubmlst.org/campylobacter/). These isolate data can be extracted using the "querying by cited publication" function.

RESULTS

Genetic diversity. MLST was completed for 1,003 isolates, of which 742 (74%) were *C. jejuni* and 261 (26%) were *C. coli* based on ST. The MLST-assigned species differed from the original speciation test with 47 (16%) *C. coli* isolates being reclassified as *C. jejuni* and 18 (2.7%) *C. jejuni* isolates reclassified as *C. coli*. There were a total of 276 STs: 192 from *C. jejuni* and 84 from *C. coli*. The 16 most common STs in each species accounted for 58% and 62% of genotypes from *C. jejuni* and *C. coli*, respectively (Fig. 1). The full breakdown by ST of this data set and the published data sets used in this paper is given in Tables S1 and S2 in the supplemental material.

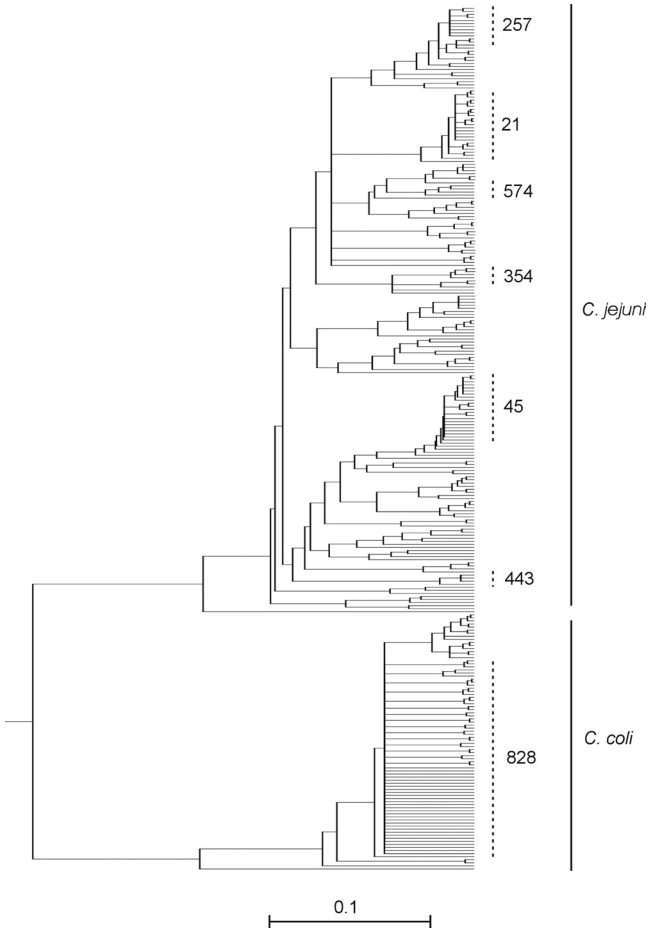

FIG. 2. Genealogy of *C. jejuni* and *C. coli* isolates from United Kingdom retail poultry samples. The tree describes a genealogy estimated by ClonalFrame using 276 STs from 1,003 typed isolates. The positions of the most abundant clonal complexes (≥5 STs) in the data set are highlighted. The scale bar is in coalescent units.

The genealogy estimated using ClonalFrame showed a high degree of genetic structuring in isolates sampled from retail poultry (Fig. 2). Annotation of the *Campylobacter* ClonalFrame tree with clonal complex designations showed that the major clonal complex groupings were supported by this phylogeny. Within *C. jejuni*, there was evidence of many recent clonal expansions corresponding to branching at the tips of the tree. The deepest-branching *C. jejuni* ST (ST-2622) contains two *C. coli* alleles. A hybrid *C. coli* ST (ST-2471) with a single *C. jejuni* allele is also present. The deepest-branching *C. coli* ST (ST-2485) corresponded to one of the three subspecies clades (clade 3) that have been previously identified in this species (38). These deep-branching subspecies *C. coli* clades differed by approximately 5% at the nucleotide level and were usually isolated from environmental waters (39). The majority of *C. coli* STs from retail poultry were from the clade previously described as clade 1 (38), which includes isolates belonging to the ST-828 clonal complex and which is the dominant clade in human disease (39).

Assignment to host. Assignment of *C. jejuni* isolates from chicken in the Netherlands ($n = 53$) and Senegal ($n = 46$) to

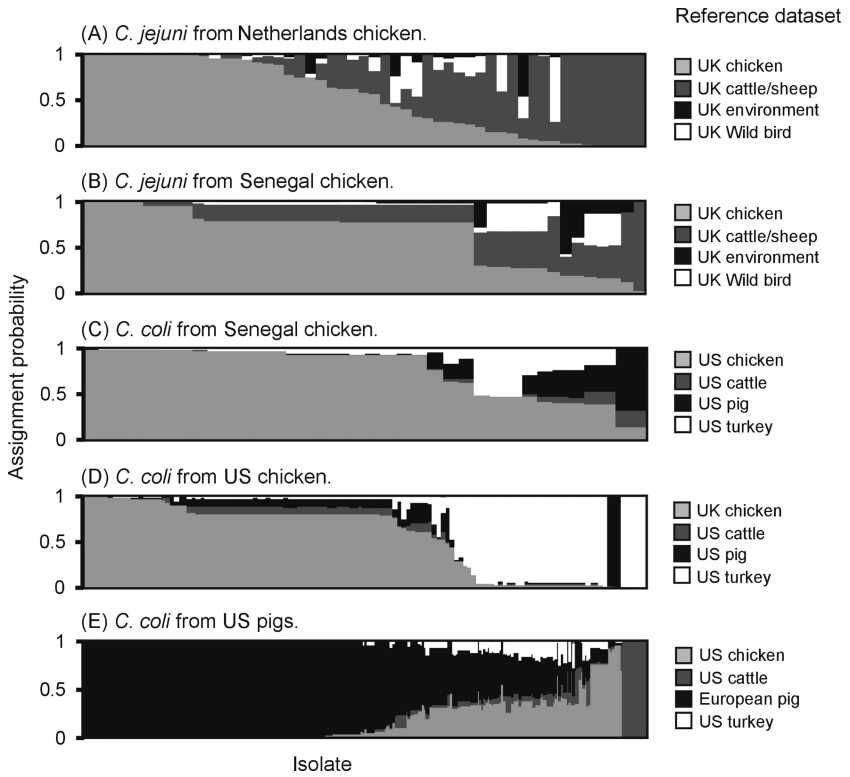

FIG. 3. Probabilistic assignment of published STs from chicken (A to D) and pig (E) isolates to host populations using the Bayesian clustering algorithm Structure. Each isolate is represented by a shaded vertical bar, showing the estimated probability that it comes from each of the putative sources. Chicken *C. jejuni* isolates from the Netherlands (A) and Senegal (B) were assigned among populations of *C. jejuni* isolates from retail poultry, ruminants (cattle and sheep), the environment, and wild birds. Chicken *C. coli* isolates from Senegal (C) were assigned among populations of chickens, cattle, pigs, and turkeys from the United States. Chicken *C. coli* isolates from the United States (D) were assigned among populations of cattle, pigs, and turkeys from the United States and retail poultry from the United Kingdom. Pig *C. coli* isolates from the United States (E) were assigned among populations of chickens, cattle, and turkeys from the United States and pig isolates from Europe.

the reference data set comprised of the current retail poultry study sample and isolates derived from ruminants, wild birds, and the environment from previously published work attributed 58% and 65% of chicken isolates, respectively, compared to 29% and 24% to the ruminant sample, with the remainder assigned between the populations of environmentally derived and wild bird isolates. This compared to the 25% to each population that would be expected if no host association signal was present in the data (Fig. 3).

For *C. coli*, Senegalese chicken isolates assigned to a reference set comprised of the U.S. isolates from chickens, cattle, pigs, and turkeys gave 76% assignment to chickens and 3%, 10%, and 11% assignments to the other populations, respectively (Fig. 3). Assignment of the U.S. populations of chicken- and pig-derived *C. coli* isolates to reference data sets, comprising isolate populations from the three other host species from the U.S. but European collections of isolates from the same host species (chickens or pigs, respectively), again produces a substantial excess of assignment to the identical host species in each case over the 25% predicted by chance. There was 55% assignment of U.S. chicken to the United Kingdom retail poultry population and there were assignments of 33%, 8%, and 4% to the sympatric turkey, swine, and cattle populations, respectively (Fig. 3). The U.S. pig isolates were assigned 72% to the European pig population and 17%, 7%, and 4% each to U.S. chicken, U.S. cattle, and U.S. turkey populations. These U.S. chicken and pig assignment results demonstrate the excess of host association over any geographical population structure that may be present.

Analysis of molecular variation (AMOVA), an alternative way to partition geographical and host-associated genetic structuring, attributed 11% of variance in haplotype sequence to host species and the rest to within-population variation for *C. jejuni* isolates from chicken and cattle isolates from the United Kingdom and Finland. No variance was attributed to country. For the *C. coli* isolates, 6.8% were attributed to host species (cattle, chicken, or pig) and again none to sampling location (United States or Europe), with the rest of the variance found between individuals within host populations.

Population differentiation as assessed by F_{ST} using the same population as in the Structure analysis produced compatible results regarding host association, demonstrating the strength of this effect (Fig. 4). Among *C. jejuni* wild bird and environmental isolates separated from food animal isolates, there was less clear separation between chicken and ruminant isolates by this approach. The results of F_{ST} analysis for *C. coli* are consistent with the Structure analysis and show the greater

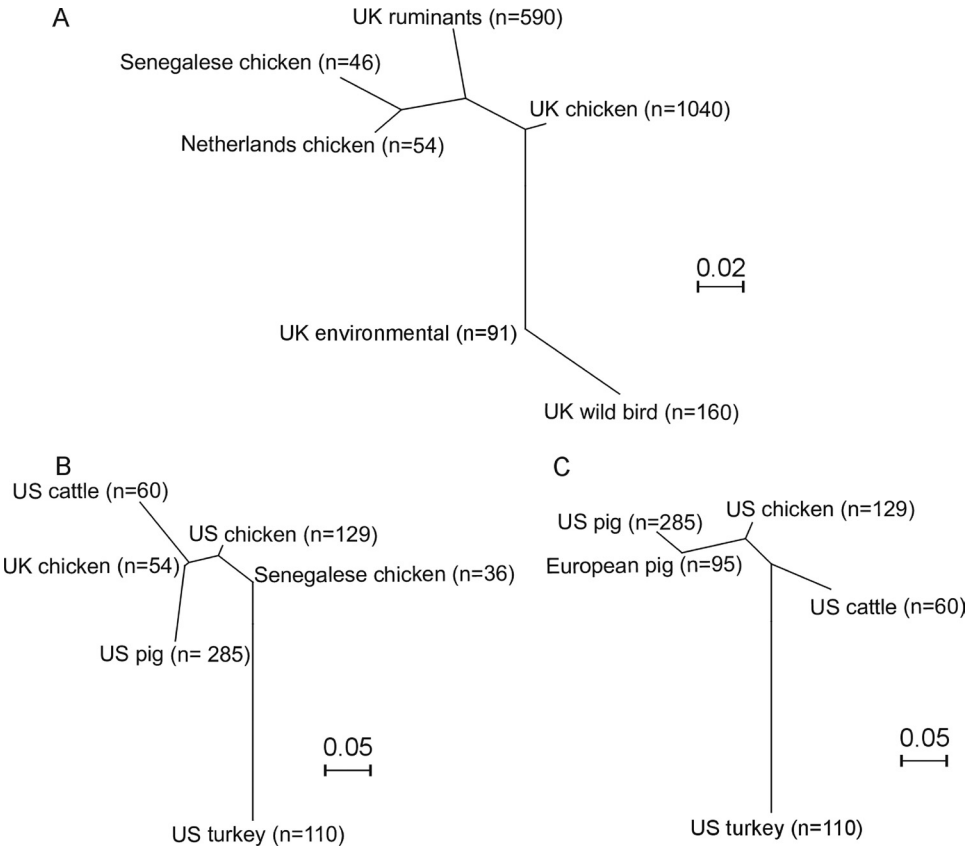

FIG. 4. F_{ST} trees based on 3,309-bp concatenated sequences of MLST genes from *C. jejuni* (A) and *C. coli* (B and C) showing genetic distance between the populations, structured by host and location of sampling, used in the Structure population genetic analysis. The scale bar is a measure of population differentiation by F_{ST} test, where 0 represents identical populations and 1 represents completely differentiated populations.

strength of host association over population structuring by geography at the levels studied.

Host-associated lineages. ClonalFrame trees were annotated with host and geographic source of isolates to explore the phylogenetic pattern of host and geographic association. These are shown for *C. jejuni* (Fig. 5) and *C. coli* (Fig. 6). There was evidence for association of hosts with recently expanded clades for both *Campylobacter* species. There is less clustering based on geography. The association between pairs of populations and clade structure was tested using permutation tests. The results of these tests are shown on each figure panel, with an arrow indicating the score from the data and a distribution of scores generated randomly from each data set (Fig. 5 and 6). The association of host with clade is well supported statistically ($P < 0.0001$) for *C. jejuni* (chickens and cattle) and *C. coli* (chickens and pigs). There was some evidence for geographic association in *C. jejuni* isolates, although the evidence was weaker than for host association. Among cattle isolates from Finland and the United Kingdom, the permutation test *P* value was 0.07. For chicken isolates, *P* values were as follows: United Kingdom versus Netherlands, $P < 0.001$; United Kingdom versus Senegal, $P < 0.001$; and United Kingdom versus Finland, $P = 0.02$. There was little evidence of geographic association with clade in *C. coli*: Europe versus United States for pig isolates, $P = 0.09$; Europe versus United States for chicken isolates, $P = 0.15$; and United Kingdom versus Senegal for chicken isolates, $P = 0.02$.

DISCUSSION

The usefulness of genetic approaches for identification of the source of human *Campylobacter* infection has become increasingly evident (33, 37, 44). This requires the development of reference data sets representing the host- and environment-associated populations from which human infections originate (9, 11, 16, 21, 23, 24, 26, 28, 29, 31, 37, 41). An understanding of the population biology of host-associated *Campylobacter* populations (29, 38), including phylogeographic effects (29), is needed to use them effectively. The practical requirements of epidemiology have similarly allowed important insights into phylogeography and global transmission networks for other bacterial pathogens (1, 22, 27).

The *Campylobacter* population associated with retail chicken meat was highly structured, with both of the principal disease-causing species, *C. jejuni* and *C. coli*, represented by multiple STs. The major lineages that were identified corresponded to clonal complex designations that have been described previously in human disease and potential sources, including chickens (5, 10, 21, 26, 28, 31, 37). This study demonstrates the diversity of *C. jejuni* isolates present on retail chicken in the

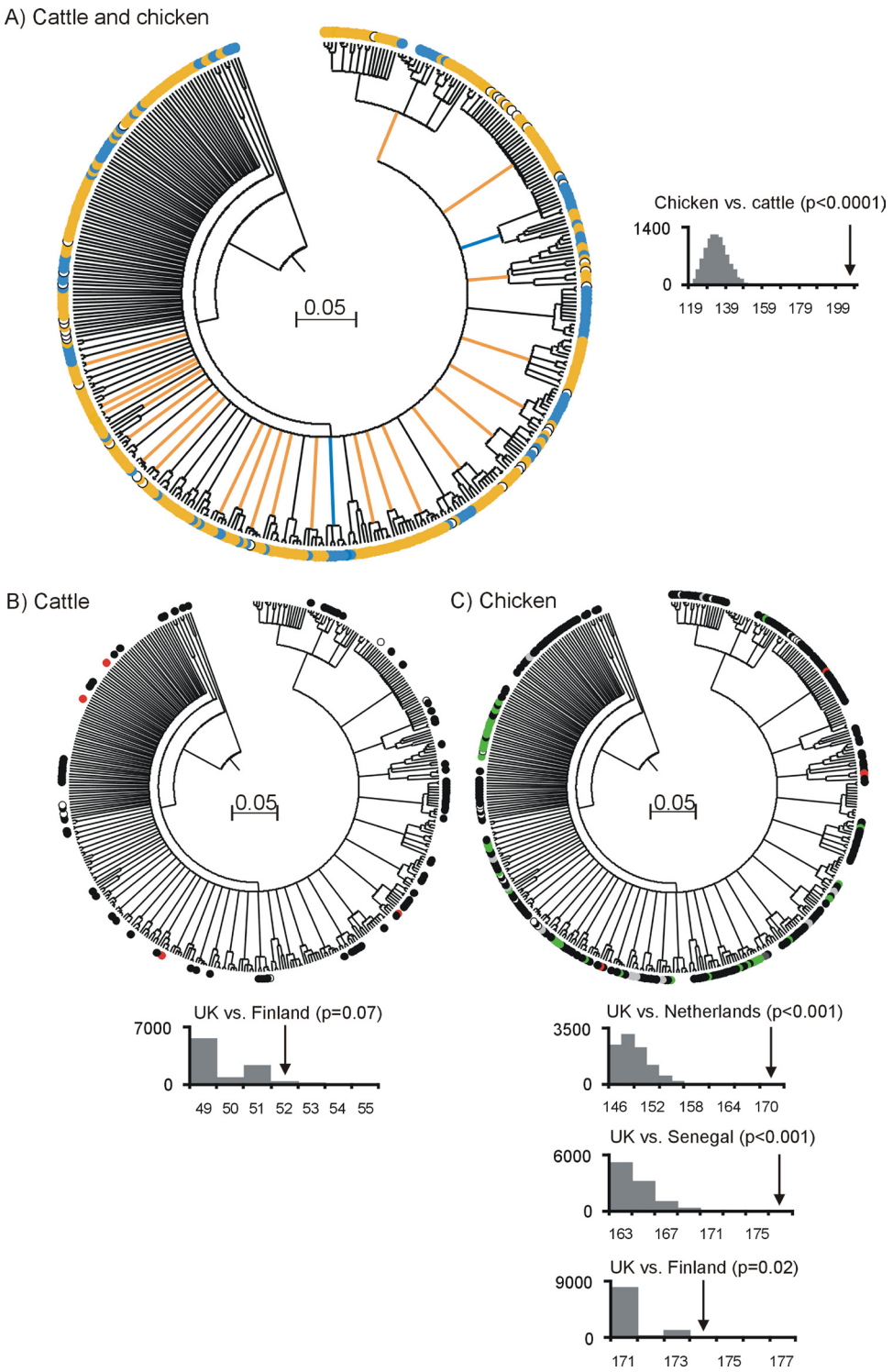

FIG. 5. ClonalFrame genealogy of *C. jejuni* isolates from retail poultry and from published data sets showing association with chicken and ruminant hosts (A) and geographic location within cattle isolates (B) and chicken isolates (C). Host source is indicated in panel A by blue (ruminants), yellow (chickens), and open (both) circles. The countries of origin are indicated in panels B and C by black (United Kingdom), red (Finland), gray (Senegal), green (Netherlands), and open (multiple countries) circles. The colored branches leading to clades in panel A indicate that the clade contains isolates, corresponding to that color, from more than one country. The scale bars are in coalescent units. Histograms in each panel show the permutation test results for the association of clade with host (A) or geographic location (B and C). The arrows show the results from the data compared to frequency histograms of the scores from 10,000 permutations of the data which show the expected distribution of scores if no association exists.

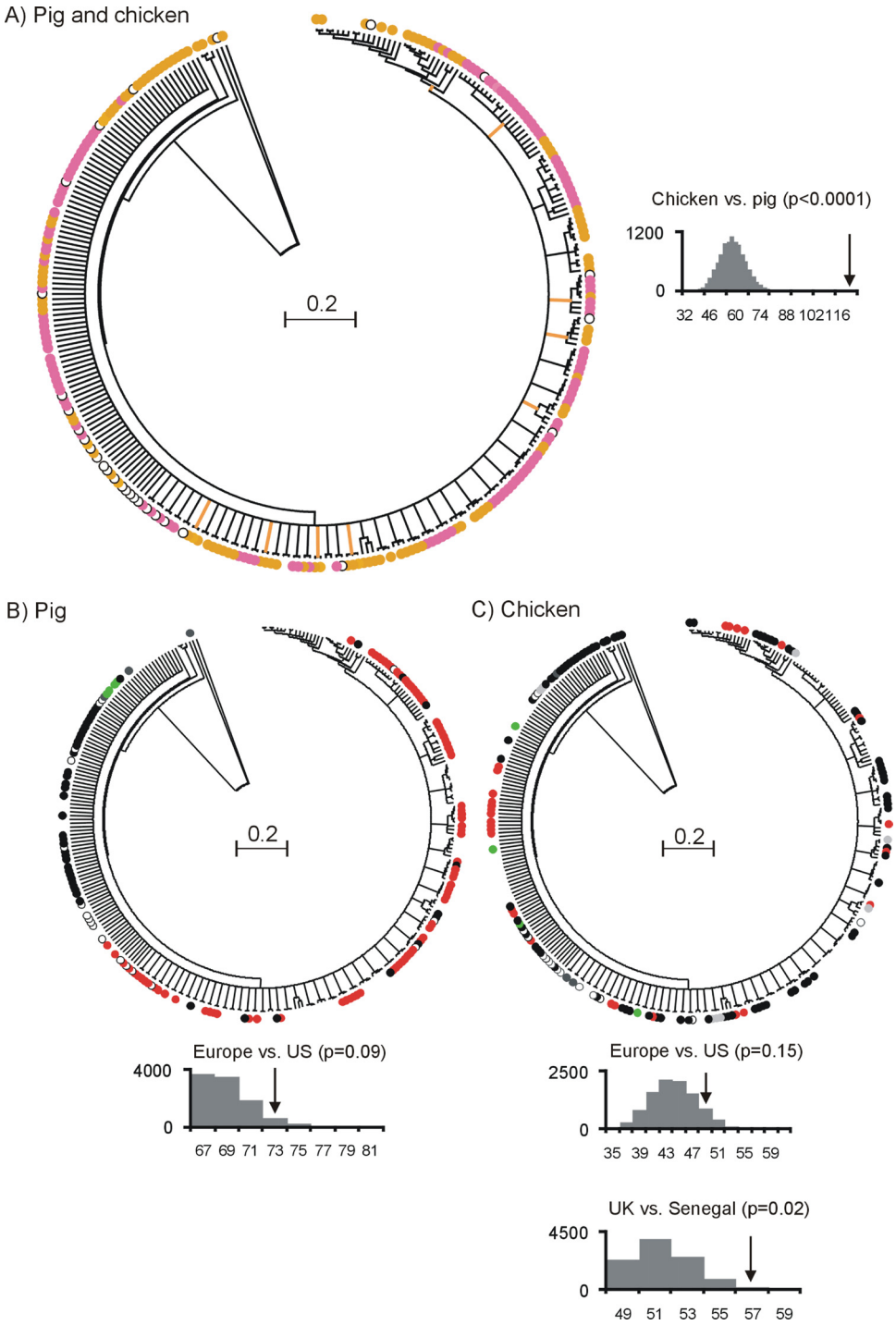

FIG. 6. ClonalFrame genealogy of *C. coli* isolates from retail poultry and from published data sets showing association with chicken and pig hosts (A) and geographic location within pig isolates (B) and chicken isolates (C). Host source is indicated in panel A by pink (pigs), yellow (chickens), and open (both) circles. The countries of origin are indicated in panels B and C by black (United Kingdom), red (United States), gray (Senegal), green (Denmark), and open (multiple countries) circles. The colored branches leading to clades in panel A indicate that the clade contains isolates, corresponding to that color, from more than one country. The scale bars are in coalescent units. Histograms in each panel show the permutation test results for the association of clade with host (A) or geographic location (B and C). The arrows show the results from the data compared to frequency histograms of the scores from 10,000 permutations of the data which show the expected distribution of scores if no association exists.

United Kingdom, including the major disease-causing lineages (37). The population of *C. coli* was dominated by isolates from one of the three previously identified subspecies clades (clade 1) (38), which includes almost all human clinical *C. coli* isolates (37, 39).

The analysis of isolates from this study together with previously published isolate collections that had been sequence typed with the same MLST scheme demonstrated strong host association within the *C. jejuni* and *C. coli* populations. This is consistent with previous work showing host-associated genetic differentiation (5, 29, 37, 44). Data from diverse hosts and geographical locations allowed the evaluation of both host association and geographic structuring on a global scale and the assessment of their relative importance. There was little evidence for geographic structure of *C. jejuni* or *C. coli*. Recently derived lineages showed some association with geography in *C. jejuni* populations, but analysis of molecular variation showed no evidence for geographical structuring in either species. The host association signals were far stronger than the phylogeographic ones. The 72% assignment of U.S. pig *C. coli* isolates to the European pig population, compared with only 28% being attributed to the three U.S. sources (chickens, cattle, and turkeys), and the 55% assignment of U.S. chicken to United Kingdom retail poultry and only 45% to U.S. sources (turkeys, pigs, and cattle) (Fig. 3) demonstrated that the host association effect substantially exceeds phylogeographic structuring even on this global scale of comparison. This may indicate the presence of a global host-associated *C. coli* gene pool in these two food animal species. The attribution of 65% of *C. jejuni* isolates from Senegalese chicken to United Kingdom retail poultry indicates that a similar host-associated gene pool may exist for *C. jejuni* in chicken. While this may be the case, global host-associated gene pools do not imply genetic homogeneity at other levels. For example, particular *C. jejuni* subtypes from chicken are strongly associated with individual production companies in New Zealand (33).

Annotation of phylogenetic trees with host source showed clustering of isolates by host with recently expanded lineages. There was strong statistical support for this in cattle and chickens (*C. jejuni*) and in pigs and chickens (*C. coli*). These host-associated lineages did not show evidence of older phylogenetic relatedness to each other. This pattern is consistent with several independent introductions and expansions within host species. The capacity for single bacterial clones to expand within the environment of the global poultry industry has been dramatically described for *Staphylococcus aureus* with the pandemic spread of a single clone entering chickens from humans in the 1970s (27). These lineage expansions within species combined with the evidence for intercontinental host attribution effects may thus indicate the capacity of food animal species to act as global transmission networks for human pathogens.

Although greater accuracy of attribution for human cases will be possible with closely matched reference populations, our findings show that precisely geographically matched reference data sets are not essential and that well-sampled reference data sets from other locations may be useful where these are not available. A balance between locally sampled populations alongside appropriately weighted wider populations will allow maximum use of available data for source attribution. There are limitations to the level of attribution accuracy that is currently possible. First, there is a lack of large well-sampled isolate collections for many possible sources. Second, the relatively large MLST collections available are restricted to seven loci, limiting the degree of genetic differentiation that can be indexed. Third, analytical approaches are imperfect. For example, the assignment approach used here allows estimation of assignment at a population level but is less useful for individual sequence types since it does not provide reliable estimates of statistical uncertainty at this level. None of these limitations are insurmountable, and the rapid development of inexpensive next generation sequencing and analytical methods to use these data will be central to facilitating the further development of these approaches to source attribution. The planned and systematic sampling of isolates from important possible sources is essential to fully exploit the opportunities that these developing technologies will provide. The United Kingdom retail poultry isolates used in this study are archived at the Health Protection Agency for such future public health applications.

This study has confirmed and extended previous findings of host association in *Campylobacter* species and in particular has shown the robustness of this association to even large geographic distances. These findings and the additional data provided can form a basis for further development of genotype-based source attribution studies to investigate the origin of human campylobacteriosis and inform targeted public health interventions.

ACKNOWLEDGMENTS

We thank Florence Opesan, Olivia Coffey, and Sophie Rollinson, Food Standards Agency—London for providing data; Keith Jolley (University of Oxford) for help in creating the database; Ella Powell, Kate Martin, Hopi Yip, and Radha Patel (Health Protection Agency, Centre for Infections) for microbiological support and data provision; David Lock (LACORS) and Ian Wilson (Northern Ireland Public Health Laboratory) for survey coordination; and staff in a wide range of participating food control laboratories (HPA, National Public Health Service—Wales and the Northern Ireland Public Health Laboratory, Public Analysts). We also thank the four anonymous referees whose comments improved the paper.

The Food Standards Agency funded genotyping and analysis. S.K.S. and M.C.J.M. are funded by Wellcome Trust Fellowships.

REFERENCES

1. **Achtman, M.** 2008. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. Annu. Rev. Microbiol. **62:**53–70.
2. **Best, E. L., N. B. Powell, C. Swift, K. A. Grant, and J. A. Frost.** 2003. Applicability of a rapid duplex real-time PCR assay for speciation of Campylobacter jejuni and Campylobacter coli directly from culture plates. FEMS Microbiol. Lett. **229:**237–241.
3. **Bolton, F. J., D. R. A. Wareing, M. B. Skirrow, and D. N. Hutchinson.** 1992. Identification and biotyping of campylobacters, p. 151–161. *In* R. G. Board, D. Jones, and F. A. Skinner (ed.), Identification methods in applied and environmental microbiology. Blackwell Scientific Publications, Ltd., London, United Kingdom.
4. **Buzby, J. C., and T. Roberts.** 1997. Economic costs and trade impacts of microbial foodbourne illness. World Health Stat. Q. **50:**57–66.
5. **Colles, F. M., T. A. Jones, N. D. McCarthy, S. K. Sheppard, A. J. Cody, K. E. Dingle, M. S. Dawkins, and M. C. Maiden.** 2008. *Campylobacter* infection of broiler chickens in a free-range environment. Environ. Microbiol. **10:**2042–2050.
6. **Corry, J. E., and H. I. Atabay.** 2001 Poultry as a source of *Campylobacter* and related organisms. Symp. Ser. Soc. Appl. Microbiol. **90:**96S–114S.
7. **Didelot, X., and D. Falush.** 2007. Inference of bacterial microevolution using multilocus sequence data. Genetics **175:**1251–1266.
8. **Didelot, X., R. Urwin, M. C. Maiden, and D. Falush.** 2009. Genealogical typing of *Neisseria meningitidis*. Microbiology **155:**3176–3186.

9. **Dingle, K. E., F. M. Colles, D. Falush, and M. C. Maiden.** 2005. Sequence typing and comparison of population biology of *Campylobacter coli* and *Campylobacter jejuni*. J. Clin. Microbiol. **43:**340–347.
10. **Dingle, K. E., F. M. Colles, R. Ure, J. Wagenaar, B. Duim, F. J. Bolton, A. J. Fox, D. R. A. Wareing, and M. C. J. Maiden.** 2002. Molecular characterisation of *Campylobacter jejuni* clones: a rational basis for epidemiological investigations. Emerg. Infect. Dis. **8:**949–955.
11. **D'lima, C. B., W. G. Miller, R. E. Mandrell, S. L. Wright, R. M. Siletzky, D. K. Carver, and S. Kathariou.** 2007. Clonal population structure and specific genotypes of multidrug-resistant *Campylobacter coli* from turkeys. Appl. Environ. Microbiol. **73:**2156–2164.
12. **Embley, T. M.** 1991. The linear PCR reaction: a simple and robust method for sequencing amplified rRNA genes. Lett. Appl. Microbiol. **13:**171–174.
13. **Excoffier, L., G. Laval, and S. Schneider.** 2005. ARLEQUIN (version 3.0): an integrated software package for population genetics data analysis. Evol. Bioinformatics Online **1:**47–50.
14. **Fearnhead, P., N. G. Smith, M. Barrigas, A. Fox, and N. French.** 2005. Analysis of recombination in *Campylobacter jejuni* from MLST population data. J. Mol. Evol. **61:**333–340.
15. **Food Standards Agency.** 27 February 2003, posting date. UK-wide survey of *Salmonella* and *Campylobacter* contamination of fresh and frozen chicken on retail sale. Food Standards Agency. http://www.food.gov.uk.
16. **French, N., M. Barrigas, P. Brown, P. Ribiero, N. Williams, H. Leatherbarrow, R. Birtles, E. Bolton, P. Fearnhead, and A. Fox.** 2005. Spatial epidemiology and natural population structure of *Campylobacter jejuni* colonizing a farmland ecosystem. Environ. Microbiol. **7:**1116–1126.
17. **Friedman, C. J., J. Neiman, H. C. Wegener, and R. V. Tauxe.** 2000. Epidemiology of *Campylobacter jejuni* infections in the United States and other industrialised nations, p. 121–138. *In* I. Nachamkin and M. J. Blaser (ed.), *Campylobacter*. ASM Press, Washington, DC.
18. **Friedman, C. R., R. M. Hoekstra, M. Samuel, R. Marcus, J. Bender, B. Shiferaw, S. Reddy, S. D. Ahuja, D. L. Helfrick, F. Hardnett, M. Carter, B. Anderson, and R. V. Tauxe.** 2004. Risk factors for sporadic *Campylobacter* infection in the United States: a case-control study in FoodNet sites. Clin. Infect. Dis. **38**(Suppl. 3)**:**S285–S296.
19. **Jolley, K. A., M. S. Chan, and M. C. Maiden.** 2004. mlstdbNet—distributed multi-locus sequence typing (MLST) databases. BMC Bioinformatics **5:**86.
20. **Kapperud, G., G. Espeland, E. Wahl, A. Walde, H. Herikstad, S. Gustavsen, I. Tveit, O. Natas, L. Bevanger, and A. Digranes.** 2003. Factors associated with increased and decreased risk of *Campylobacter* infection: a prospective case-control study in Norway. Am. J. Epidemiol. **158:**234–242.
21. **Karenlampi, R., H. Rautelin, D. Schonberg-Norio, L. Paulin, and M. L. Hanninen.** 2007. Longitudinal study of Finnish *Campylobacter jejuni* and *C. coli* isolates from humans, using multilocus sequence typing, including comparison with epidemiological data and isolates from poultry and cattle. Appl. Environ. Microbiol. **73:**148–155.
22. **Keim, P. S., and D. M. Wagner.** 2009. Humans and evolutionary and ecological forces shaped the phylogeography of recently emerged diseases. Nat. Rev. Microbiol. **7:**813–821.
23. **Kinana, A. D., E. Cardinale, I. Bahsoun, F. Tall, J. M. Sire, S. Breurec, B. Garin, C. Saad-Bouh Boye, and J. D. Perrier-Gros-Claude.** 2007. *Campylobacter coli* isolates derived from chickens in Senegal: diversity, genetic exchange with *Campylobacter jejuni* and quinolone resistance. Res. Microbiol. **158:**138–142.
24. **Kinana, A. D., E. Cardinale, F. Tall, I. Bahsoun, J. M. Sire, B. Garin, S. Breurec, C. S. Boye, and J. D. Perrier-Gros-Claude.** 2006. Genetic diversity and quinolone resistance in *Campylobacter jejuni* isolates from poultry in Senegal. Appl. Environ. Microbiol. **72:**3309–3313.
25. **Kumar, S., K. Tamura, and M. Nei.** 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinform. **5:**150–163.
26. **Litrup, E., M. Torpdahl, and E. M. Nielsen.** 2007. Multilocus sequence typing performed on *Campylobacter coli* isolates from humans, broilers, pigs and cattle originating in Denmark. J. Appl. Microbiol. **103:**210–218.
27. **Lowder, B. V., C. M. Guinane, N. L. Ben Zakour, L. A. Weinert, A. Conway-Morris, R. A. Cartwright, A. J. Simpson, A. Rambaut, U. Nubel, and J. R. Fitzegerald.** 2009. Recent human-to-poultry host jump, adaptation, and pandemic spread of Staphylococcus aureus. Proc. Natl. Acad. Sci. U. S. A. **106:**19545–19550.
28. **Manning, G., C. G. Dowson, M. C. Bagnall, I. H. Ahmed, M. West, and D. G. Newell.** 2003. Multilocus sequence typing for comparison of veterinary and human isolates of *Campylobacter jejuni*. Appl. Environ. Microbiol. **69:**6370–6379.
29. **McCarthy, N. D., F. M. Colles, K. E. Dingle, M. C. Bagnall, G. Manning, M. C. Maiden, and D. Falush.** 2007. Host-associated genetic import in *Campylobacter jejuni*. Emerg. Infect. Dis. **13:**267–272.
30. **Meldrum, R. J., D. Tucker, and C. Edwards.** 2004. Baseline rates of Campylobacter and Salmonella in raw chicken in Wales, UK in 2002. J. Food Prot. **67:**1226–1228.
31. **Miller, W. G., M. D. Englen, S. Kathariou, I. V. Wesley, G. Wang, L. Pittenger-Alley, R. M. Siletz, W. Muraoka, P. J. Fedorka-Cray, and R. E. Mandrell.** 2006. Identification of host-associated alleles by multilocus sequence typing of *Campylobacter coli* strains from food animals. Microbiology **152:**245–255.
32. **Miller, W. G., S. L. On, G. Wang, S. Fontanoz, A. J. Lastovica, and R. E. Mandrell.** 2005. Extended multilocus sequence typing system for *Campylobacter coli*, *C. lari*, *C. upsaliensis*, and *C. helveticus*. J. Clin. Microbiol. **43:**2315–2329.
33. **Mullner, P., S. E. Spencer, D. J. Wilson, G. Jones, A. D. Noble, A. C. Midwinter, J. M. Collins-Emerson, P. Carter, S. Hathaway, and N. P. French.** 2009. Assigning the source of human campylobacteriosis in New Zealand: a comparative genetic and epidemiological approach. Infect. Genet. Evol. **9:**1311–1319.
34. **Neimann, J., J. Engberg, K. Molbak, and H. C. Wegener.** 2003. A case-control study of risk factors for sporadic *Campylobacter* infections in Denmark. Epidemiol. Infect. **130:**353–366.
35. **Pritchard, J. K., M. Stephens, and P. Donnelly.** 2000. Inference of population structure using multilocus genotype data. Genetics **155:**945–959.
36. **Sheppard, S. K., J. F. Dallas, M. MacRae, N. D. McCarthy, E. L. Sproston, F. J. Gormley, N. J. Strachan, I. D. Ogden, M. C. Maiden, and K. J. Forbes.** 2009. Campylobacter genotypes from food animals, environmental sources and clinical disease in Scotland 2005/6. Int. J. Food Microbiol. **134:**96–103.
37. **Sheppard, S. K., J. F. Dallas, N. J. Strachan, M. MacRae, N. D. McCarthy, D. J. Wilson, F. J. Gormley, D. Falush, I. D. Ogden, M. C. Maiden, and K. J. Forbes.** 2009. Campylobacter genotyping to determine the source of human infection. Clin. Infect. Dis. **48:**1072–1078.
38. **Sheppard, S. K., N. D. McCarthy, D. Falush, and M. C. Maiden.** 2008. Convergence of *Campylobacter* species: implications for bacterial evolution. Science **320:**237–239.
39. **Sopwith, W., A. Birtles, M. Matthews, A. Fox, S. Gee, S. James, J. Kempster, M. Painter, V. Edwards-Jones, K. Osborn, M. Regan, Q. Syed, and E. Bolton.** 2010. Investigation of food and environmental exposures relating to epidemiology of *Campylobacter coli* in humans in northwest England. Appl. Environ. Microbiol. **76:**129–135.
40. **Staden, R.** 1996. The Staden sequence analysis package. Mol. Biotechnol. **5:**233–241.
41. **Thakur, S., and W. A. Gebreyes.** 2005. *Campylobacter coli* in swine production: antimicrobial resistance mechanisms and molecular epidemiology. J. Clin. Microbiol. **43:**5705–5714.
42. **Vellinga, A., and F. Van Loock.** 2002. The dioxin crisis as experiment to determine poultry-related *Campylobacter* enteritis. Emerg. Infect. Dis. **8:**19–22.
43. **Wilson, D. J., E. Gabriel, A. J. Leatherbarrow, J. Cheesbrough, S. Gee, E. Bolton, A. Fox, C. A. Hart, P. J. Diggle, and P. Fearnhead.** 2009. Rapid evolution and the importance of recombination to the gastroenteric pathogen Campylobacter jejuni. Mol. Biol. Evol. **26:**385–397.
44. **Wilson, D. J., E. Gabriel, A. J. H. Leatherbarrow, J. Cheesbrough, S. Gee, E. Bolton, A. Fox, P. Fearnhead, A. Hart, and P. J. Diggle.** 2008. Tracing the source of campylobacteriosis. PLoS Genet. **26:**e1000203.